The Monomerization of the Purple Protein, A Member of the GFP-family.

Brooke Corning

The University of Texas at Austin
Dean's Scholars Honors Thesis
2009

Supervising Professor: Scott Stevens, Ph.D.

_____
Signature Date

Honors Advisor: Ruth Buskirk, Ph.D.

_____
Signature Date

Table of Contents

Abstract

Author: Brooke Corning
Title: The Monomerization of the Purple Protein, A Member of the GFP-family
Supervising Professor: Scott Stevens, Ph.D.

Green fluorescent protein (GFP) has been used extensively since its discovery in the 1960s to report and visualize gene expression. For years it has been the only known naturally occurring fluorescent pigment that is encoded by a single gene, making it extremely useful in various fields of biology, because the expression of this gene directly leads to the appearance of the fluorescent green color. Recently, however, many more proteins with similar properties to GFP, and available in a variety of colors, have been isolated from the class of marine organisms called Anthozoa, which includes the corals. This increase in the availability of colored proteins in the GFP family in turn has expanded the number of available biotechnology applications. However, some of these newly discovered GFP-like proteins do not have wild-type forms that readily allow for the creation of fusion proteins, particularly because of oligomerization. It is widely accepted that almost all members of the GFP-family form dimers or tetramers in their functional forms. This study investigates a GFP-like protein, Purple, isolated from two species, *Galaxea fascicularis* and *Montipora efflorescens*. Purple protein forms oligomers when expressed, which would then interfere with the normal expression of a protein to be tagged in gene fusion experiments. We selectively mutated 3 amino acids, which we believed were responsible for oligomerization in Purple. These 3 residues were chosen based on sequence similarities to a very similar protein, a mutant form of the Rtms5 chromoprotein from *Montipora efflorescens*. While we had hoped that the resulting triple-mutant Purple protein would form monomers in vivo while retaining its

purple coloration, this turned out to be incorrect. The resulting mutants had lost their ability to turn purple. However, we also determined that we had successfully changed the oligomerization state of Purple by examining the relative molecular mass of one our mutant proteins, which turned out to be half the size of the original purple protein. It is possible that by adding addition mutations in the future, the original spectral properties could be recovered. If successful, this would further expand the utility of the GFP family.

Background

Green Fluorescent Protein (GFP) has become one of the most widely studied and used proteins in biology today. The protein was originally isolated from the *Aequora* jellyfish in 1962 by Osama Shimimorua as a companion protein to aquorin (Tsien, 1998). However, it subsequently took over forty years for the true value of GFP to be fully realized, and it was not until 2008 that the Nobel Prize in chemistry was awarded to Osamu Shimomura, Martin Chalfie, and Roger Tsien for the discovery and development of GFP. After the discovery of the protein itself, one of the most important subsequent findings was that the gene could also be expressed in organisms other than *Aequora* and still maintain its fluorescence (Tsien, 1998). The significance of this was that the gene contained all the necessary information for the synthesis of the chromophore without requiring the addition of other enzymes or substrates as is the case with fluorescent jellyfish proteins. Several other genera of marine organisms related to the jellyfish were found to have GFP-like proteins, but all of these seemed to have chromophores that were external cofactors, such as lumazines or flavins (Tsien, 1998).

Structure of GFP

Wild-type GFP is a stable protein consisting of a single chain of 238 residues. It has absorption maxima at both 395 and 475 nm (Ormo, 1996), and an emission peak at 475 nm (Yang, 1996). The chromophore is a *p*-hydroxybenzylideneimidazolinone, which is formed by residues 65–67: Serine, Tyrosine, and Glycine, respectively, in the protein (Tsien, 1998). The chromophore is formed by autocatylitic cyclicization and oxidation of these three amino acids, and excitation is then triggered with a specific wavelength of light (Phillips, 2001).

After determining the crystal structure of GFP, it was discovered that that the protein fold consists of an 11-stranded beta-barrel, with a coaxial helix in the middle. The chromophore forms the central helix (Ormo, 1996). Additionally, there are short helical segments at the end of the cylinder formed by the beta-barrel, which form caps. The fluorophore is protected from the outside environment on the central helix within the beta barrel. This barrel with caps essentially forms a can, which does not have any holes though which ligands could diffuse through. This structure could explain the stability of GFP, including characteristics such as resistance to heat and other denaturants (Yang, 1996). Wild-type GFP folds efficiently around room temperature, but this efficiency rapidly declines at higher temperatures. However, once it is folded properly, it can fluoresce at higher temperatures. Subsequent variants produced by mutations have increased the temperature at which GFP is able to fold (Tsien, 1998).

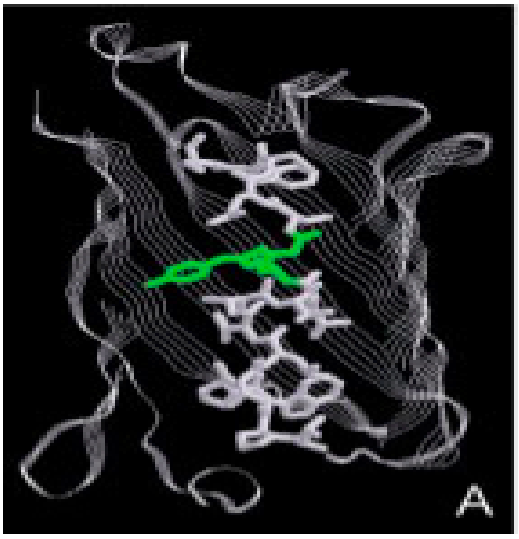


Fig 1. Barrel structure of GFP. The chromophore is colored green. Adapted from: Matz "Family of the Green Fluorescent Protein: Journey to the End of the Rainbow." (2002).

Modifications to the Original GFP

The original (wild-type) version of GFP that was isolated from *Aqueora* had a number of problems that had to be corrected to optimize its use in a variety of applications. The formation of the fluorophore was slow, it often formed inactive inclusion bodies in the cell, and the fluorescent intensity was low (Phillips, 2001). Also previously noted were the problems GPF had when attempting to fold at higher temperatures. Thus, attempts were made to modify wild-type GFP to correct these problems. It was found that a single amino acid substitution could increase the speed of the fluorophore formation from 2 hours to 45 minutes, and additional mutations allowed the protein to fold more efficiently at higher temperatures. Mutants were also created in which two absorbance peaks in wild-type GFP were converted to a single absorbance peak. In combination, these mutations were used to create a version of GFP known as enhanced GFP (EGFP), characterized by enhanced stability and brightness. EGFP can be visualized in cells at low intensity for several hours with minimal photobleaching (Lippincott-Schwartz, 2003). Mutants were also screened for a variety of other desirable properties, particularly increased brightness.

Another valuable characteristic in the GFP mutants was color variation. During the nineties, in the absence of the discovery of additional, non-green proteins with the same properties as GFP, an effort was made to create other colors from the original green-fluorescent protein isolated from *Aequorea victoria*. This resulted in a group of red-shifted mutants (Delagrave, 1994). These mutants were created using combinatorial mutagenesis and were screened using digital imaging spectroscopy, which allows for

the simultaneous screening of thousands of colonies. These red-shifted mutants have excitation maxima about 100 nm higher (490nm) than wild-type GFP (390nm). This difference in excitation maxima means that both of these proteins could be simultaneously expressed in the same cell, and still be spectrally separable (Delagrave, 1994). Mutational analysis of GFP has created several different isoforms in addition to the red-shifted mutants, including proteins that emit a blue-green light, a red-shifted variant, and a third variant, which is brighter than the wild-type (Misteli, 1997). These proteins were known as blue fluorescent protein (BFP), yellow-fluorescent protein (YFP), and cyan fluorescent protein (CFP).

However, none of these GFP variants were able to achieve an emission maxima longer than 529nm (Baird, 2000). There were also other shortcomings that made these mutant GFP-variants less than ideal. For example, the blue variants created in this way were dim and photobleached easily (Lippincott-Schwartz, 2003). So, despite these attempts to create color variants of GFP, one of the major limits to developing new technology has been the limited number of colors available. At best, the color variations created by mutants of GFP have allowed for dual-color labeling in experiments where it is necessary to simultaneously track two different proteins. However, there were not enough color differences to allow for triple-color labeling. The limited number of color variants also hinders the use of applications based on fluorescence resonance energy transfer (FRET). FRET technology allows for the study of protein-protein interactions in situ (Matz, 2002).

Applications of GFP

GFP is useful for a variety of applications because of its ability to fluoresce when fused to a target protein without the addition of any exogenous substrates. Thus it is often used as a reporter gene for studying protein and DNA localization. Use of GFP as a reporter gene enables the study of such things as DNA replication and translation, protein export, and signal transduction in living cells. This can be done in real time, allowing the visualization of protein localization (Phillips, 2001).

Other traditional methods of studying protein localization include such techniques as immunofluorescent microscopy. However, this requires the use of antibodies directed against the target protein, which may be difficult to obtain. Additionally, this method often requires the cells to be fixed, and thus living cells can't be used. Gene fusions using GFP have significant advantages over this method. They are relatively easy to construct, have a high level of sensitivity, and do not require invasive sample preparation (which would kill the cells) or addition of a substrate to activate the fluorescence (Phillips, 2001). This tool has led to significant discoveries in several fields. For example, as a result of the use of GFP fusion studies in bacteria, the previously held concept that bacterial cells had limited cellular organization was abandoned in light of the knowledge that there is considerable amount of sequestration of molecules into subcellular compartments (Phillips, 2001).

Fluorescent GFP has also been expressed in a number of different organisms besides bacteria. These include yeast, slime molds, plants, Drosophila, zebrafish, and

mammalian cells (Yang, 1996). In fact, one of the first uses of GFP in time-lapse microscopy tracked ribonucleoprotein particles moving through developing egg chambers of drosophila (Misteli, 1997). The ability to create a fusion protein with GFP allows it to be used as cell lineage tracers, reporters of gene expression, and a measure of protein-protein interactions (Yang, 1996). Additionally, GFP has great potential for use in biological screening. In simple cases, GFP can be fused to the gene of interest, and then used to rapidly screen cells to see if they have incorporated the gene. Other examples include more complex cases in which GFP has been used to screen embryonic stem cells and transgenic animals. Several studies have already used GFP to screen mouse embryos, resulting in a high yield of transgenic mice with low toxicity (Misteli, 1997).

The creation of gene fusions with GFP is relatively simple. A variety of commercially available vectors are available with different GFP variants. The general strategy for constructing GFP fusion proteins is to engineer restriction enzymes sites into both GFP and the target protein DNA sequence to create a recombinant plasmid. The GFP reporter gene can be added to either end of the target gene. However, if GFP is being attached to a membrane protein, GFP must be added to the cytoplasmic side (Phillips, 2001). It must be noted that the fusion of GFP to a target protein can alter both the protein's function and localization, although there are many cases where it does neither (Phillips, 2001).

The Search for GFP-like Proteins

Since the discovery of GFP in the early 1960s, the search has been on for other proteins like it. Recall that GFP is unique in its ability to produce green color from a single gene without the help of additional enzymes and cofactors. Thus, there was much interest in discovering other proteins with this property. Since GFP had originally been identified in the bioluminescent jellyfish *Aequorea victoria*, scientists naturally searched for similar proteins in similar (bioluminescent) organisms. However, many other bioluminescent organisms that were examined, such as other jellyfish and copepods, were not homologous to *Aequorea*, and in fact had evolved recently from unrelated lineages (Matz, 2002). Thus, GFP remained the only known protein to possess this unique ability until 1999. It wasn't until the search parameters were modified that other members of the GFP family started to turn up (See Fig 3). The idea emerged that GFP-like proteins might not all be fluorescent, but might simply have a high ability to absorb light (Matz, 2002). This new criteria led researchers to examine non-florescent organisms, such as the *Anthozoa* class containing brightly colored coral. And thus, new proteins, similar to GFP, started to appear everywhere.

To date, there have been over 30 significantly different GFP-like proteins that have been cloned and added to the sequence database. Of these, 6 are from bioluminescent organisms. The rest are all from *Anthozoa*, and are fluorescent but not bioluminescent (Matz, 2002). These proteins can be divided into 4 main color classes: green, yellow, orange-red, and non-fluorescent purple-blue (Alieva, 2008). All of these classes have

maintained the same structure as the original GFP isolated from *Aequorea victoria*, consisting of a beta-barrel (See Fig 2).

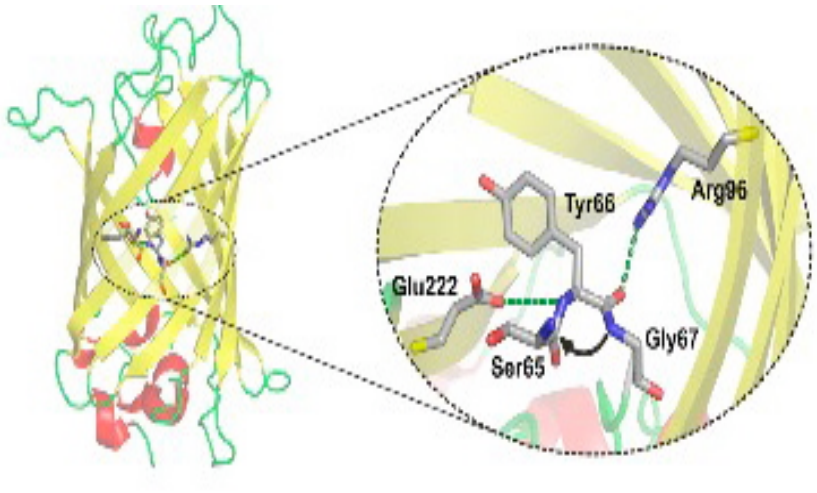
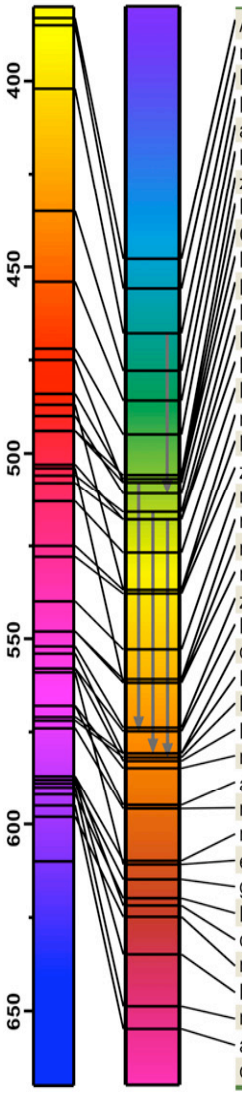


Fig 2. The beta-barrel structure in GFP-like proteins, and the pre-cyclization state of the catalytic residues. Adapted from Pakhomov, "GFP Family: Structural Insights into Spectral Tuning." (2008.)

In addition to the discovery of fluorescent, but non-bioluminescent, GFP-like proteins, it has also been shown that there are non-fluorescent GFP-like proteins that control coloration. Such an example is seen in the sea anenome *Anemonia sulcata*, which expresses a purple, non-fluorescent, GFP-like homolog in the tips of its tentacles (Lukyanov, 2000). The discovery of GFP-like proteins available in a multitude of colors has made it possible to greatly expand the previously developed technology, as well as develop new technologies for use with such proteins. The availability of red GFP-like proteins alone has improved the fluorescence in animal tissues, which typically is greatly reduced when using longer wavelength colors (Matz, 2002). Also, the selection of a variety of different colors allows for the use of multi-color labeling that was previously limited to, at most, 2 colors. It is now possible to use up to 4 colors simultaneously.



Protein	Absorbance maximum, nm	Fluorescence emission maximum, nm	Molar extinction coefficient, $x10^3$ $M^{-1}cm^{-1}$	QY[a]	Subfamily/ Chromophore	References[b]
Azurite	384	450	26.2	0.55	GFP-Y66H	Mena et al., 2006
mKalama1	385	456	36	0.45	GFP	Ai et al., 2007
PS-CFP (cyan)[c]	402	468	34	0.16		Chudakov et al., 2004; Lukyanov et al., 2005
ECFP	434	477	26	0.4	GFP-Y66W	Heim et al., 1994; Patterson et al., 2001
amFP486	454	486	ND[f]	0.71	GFP	Matz et al., 1999; Henderson and Remington, 2005
MiCy	472	495	27.25	0.9		Karasawa et al., 2004
zFP506	494	506	ND	ND	GFP	Matz et al., 1999; Pakhomov and Martynov, 2007
EGFP	484	507	53	0.6	GFP	Heim et al., 1995; Patterson et al., 1997
GFP	475; 395	508	25	0.79	GFP	Ormo et al., 1996; Patterson et al., 1997
DendFP (green)[c]	492	508	90	0.65	GFP	Pakhomov et al., 2004; Gurskaya et al., 2006
PS-CFP (green)[c]	490	511	27	0.19		Chudakov et al., 2004; Lukyanov et al., 2005
EosFP (green)[c]	506	516	72	0.70	GFP	Wiedenmann et al., 2004; Nienhaus et al., 2005
Dronpa	503	518	95	0.85	GFP	Ando et al., 2004; Andresen et al., 2007
Kaede (green)[c]	508	518	98.8	0.80	GFP	Ando et al., 2002; Mizuno et al., 2003
EYFP	514	527	84	0.61	GFP	Ormo et al., 1996; Wachter et al., 1998
mHoneydew[d]	487	537	17	0.12		Shaner et al., 2004
PhiYFP	525	537	115	0.60		Shagin et al., 2004
zFP538	528	538	ND	ND	DsRed-related	Matz et al., 1999; Remington et al., 2005
mBanana	540	553	6	0.70		Shaner et al., 2004
mKO	548	561	51.6	0.60		Karasawa et al., 2004
mHoneydew[d]	504	562	17	0.12		Shaner et al., 2004
mOrange	548	562	71	0.69	DsRed-related	Shaner et al., 2004; Shu et al., 2006
z2FP574	552	574	ND	ND	DsRed-related	Labas et al., 2002; Pakhomov and Martynov, 2007
DendFP (red)[c]	557	575	35	0.68	Kaede	Pakhomov et al., 2004; Gurskaya et al., 2006
dTomato	554	581	69	0.69		Shaner et al., 2004
EosFP (red)[c]	571	581	41	0.55	Kaede	Wiedenmann et al., 2004; Nienhaus et al., 2005
Kaede (red)[c]	572	582	60.4	0.33	Kaede	Ando et al., 2002; Mizuno et al., 2003
DsRed	558	583	75	0.79	DsRed	Matz et al.,1999; Gross et al.,2000; Shaner et al.,2004
mTangerine	568	585	38	0.30		Shaner et al., 2004
asFP595 (KFP)	568	595	56.2	<0.001	DsRed-related	Lukyanov et al., 2000; Tretyakova et al., 2007
mStrawberry	574	596	90	0.29	DsRed	Shaner et al., 2004; Shu et al., 2006
mCherry	587	610	72	0.22	DsRed	Shaner et al., 2004; Shu et al., 2006
eqFP611	559	611	78	0.45	DsRed	Wiedenmann et al., 2002; Petersen et al., 2003
gtCP[e]	589	615	ND	chromo	DsRed	Gurskaya et al., 2001; Martynov et al., 2003
Rtms5[e]	592	620	80	chromo	DsRed	Prescott et al., 2003
cgCP[e]	587	622	ND	chromo	DsRed	Gurskaya et al., 2001; Pakhomov et al., 2006
mRaspberry	598	625	86	0.15		Wang et al., 2004
Katushka	588	635	65	0.34		Shcherbo et al., 2007
mPlum	590	649	41	0.10		Wang et al., 2004
aeCP-AQ143	595	655	90	0.04		Shkrob et al., 2005
cjBlue	610	ND	66.7	chromo	DsRed	Chan et al., 2006

Fig 3. Spectral Characteristics of the GFP Family. Figure borrowed from Pakhomov, "GFP Family: Structural Insights into Spectral Tuning." (2008.)

Limitations Associated with GFP-like Proteins

While a variety of GFP-like proteins have been, and are being, discovered, there are a number of different requirements that are necessary for these proteins to be useful in biotechnology applications. Additionally, different types of experiments will need different types of proteins. Thus, while it may at first be unclear as to why such a large

variety of GFP-like proteins are desirable, it becomes more apparent when the different parameters of the experiment are examined. First, the GFP-like protein must not be toxic to the cells it is expressed in, and it must be bright enough to be detected and imaged. The brightness of a GFP-like protein is determined by several factors, including the extinction coefficent and quantum yield. Additionally, the imaging equipment used must be a good match for a given protein. The protein must also be stable, and not interfere with other cellular events (Shaner, 2005). And, as mentioned above, the GFP-like protein chosen for fusion experiments must not oligomerize. In addition to being problematic in fusion proteins, tetrameric proteins have often been found to be toxic to bacteria, especially in cases when there is a significant amount of aggregation. However, monomeric fluorescent proteins are usually not toxic to cells (Shaner, 2005). Another consideration is the temperature range in which the protein is able to mature. Many experiments are conducted at 37°C, but this is not the optimal maturation temperature for many GFP-like proteins. Photostability is another important consideration in choosing a fluorescent protein, because there is a significant amount of photobleaching between otherwise very similar proteins. This is a particularly important characteristic when an experiment requires a large number of pictures to be taken (Shaner, 2005). Currently, one of the best monomers available in the far-red (purple) spectral class, in terms of brightness and photostability, is mPlum. However, it is significantly less bright than many of the shorter-wavelength proteins (Shaner, 2005).

Considering these requirements, many of the newly discovered GFP-like proteins are not without their limitations. Most of them have not been isolated in a “readily usable

state" because they form oligomers when expressed. Recall that an important requirement when choosing a fluorescent protein for use as a fusion protein is that the fluorescent protein must not oligomerize (Shaner, 2005). This is a significant limitation because it both hinders their usefulness in FRET technology as well labeling proteins in situ (Matz, 2002). However, GFP-like proteins that form oligomers are still useful as reporter genes, selection markers, and biosensors (Shaner, 2005). Additionally, while a number of different-colored proteins have been discovered, most of these center around two regions of the visible light spectrum: cyan-green (470-520 nm), and orange-red (570-600 nm). This leaves several notably large gaps including blue-purple and yellow-orange. However, these gaps are expected to become smaller and smaller as more coral proteins are screened and studied (Matz, 2002).

A particularly good example is DS Red, which is a GFP-like fluorescent protein isolated from the genus of coral *Discosoma*. It has generated excitement as a fusion protein because of its high extinction coefficient, as well as its resistance to extreme pH and photobleaching. Additionally, its natural emission maxima of 583nm can be shifted to 602nm by a single point-mutation. This is significant because 583nm is the longest reported wavelength for a wild-type, spontaneously fluorescent protein. However, it is severely limited in its use in biotechnology applications by its strong tendency to oligomerize into tetramers and slow maturation rate (Baird, 2000).

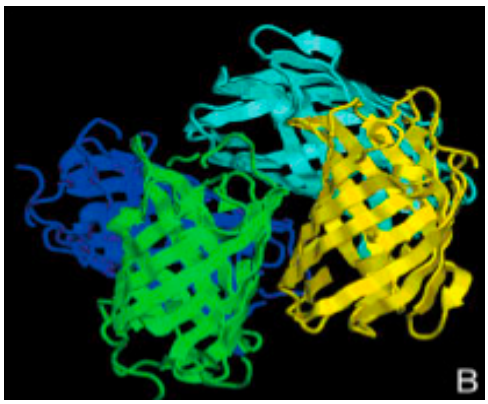


Fig 4. Tetrameric form of DsRed. Adapted from: Matz "Family of the Green Fluorescent Protein: Journey to the End of the Rainbow." (2002).

Focus of This Study: Purple Protein

The protein I am working on, known as "Purple", has been isolated from *Galaxea fascicularis* and *Montipora efflorescens*. However, Purple forms tetramers in vivo, which is an undesirable property when using the protein in fusion experiments. Our goal was to monomerize purple via site-directed mutagenesis of 3 amino acids: #102 (Threonine to Lysine), #149 (Arginine to Glutamic Acid), and #158 (Phenylalanine to Lysine). Because neither the crystal or solution structure of Purple has been determined, we depended on sequence data to compare Purple with another similar GFP-like protein, in which the 3-dimensional structure had been analyzed, and the protein had been successfully converted from a tetramer to a monomer. A mutant form of the Rtms5 chromoprotein from *Montipora efflorescens* (see Fig 5) was chosen for this purpose. Sequence comparisons indicated which of the amino acids in Purple we would most likely need to mutate based on homology to the 3 amino acids, which are responsible for allowing the protein to form oligomers (see Fig 6). Amino acid 149 of purple was chosen because in the homologous residue of Rtms5, this amino acid forms a salt bridge between subunits of the tetramers. Amino acids 102 and 158 were chosen for similar reasons.

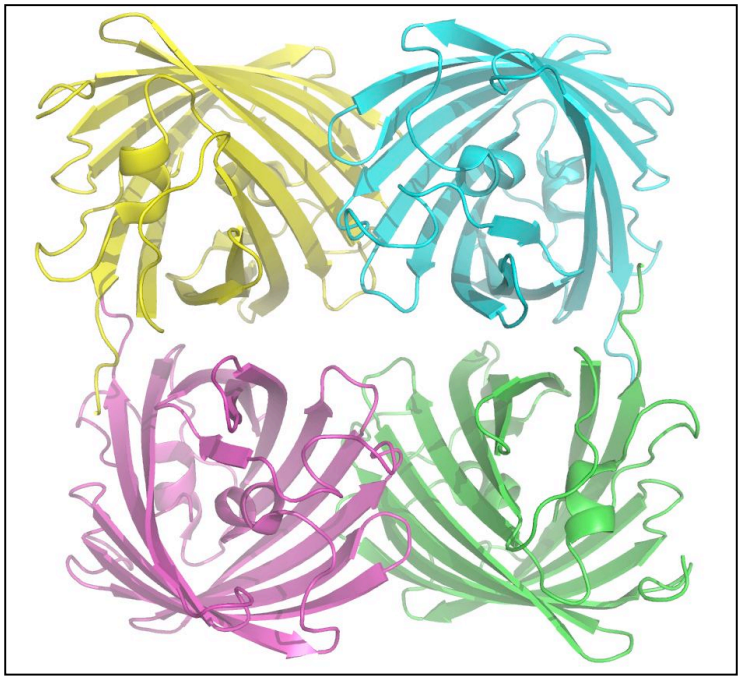


Fig 5. Structure of mutant form of the Rtms5 chromoprotein from Montipora efflorescens. Figure created by Eric J. Montemayor (University of Texas at Austin).

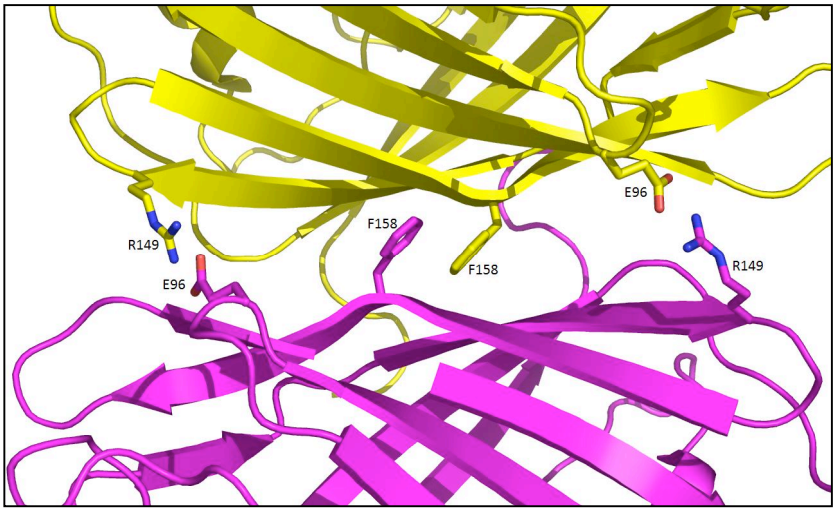


Fig 6. Interaction of residues to allow oligomerization in Purple. . Figure created by Eric J. Montemayor (University of Texas at Austin).

In order to change these three amino acids via site-directed mutagenesis, the genes for the purple proteins were inserted into a plasmid, which then allowed us to mutagenize in the bacterial overexpression plasmid. We then sequenced each mutant to confirm that

the targeted amino acid had been appropriately changed and that the rest of the gene was not mutagenized. In this way, we first generated a set of 3 single-mutants, which we then used to produce a set of 3 double-mutants. Finally, we used the double mutants to produce a single triple-mutant.

After generating the triple mutant of Purple, we assessed the resulting protein to determine whether it was still being expressed, and if it was soluble. Additionally, we then had to determine whether we had successfully created a monomer and if the mutant had maintained its ability to turn purple. To determine if our mutants had indeed been converted to a monomeric form, we purified the mutants and used size exclusion chromatography to determine its relative molecular weight. It was even simpler to ascertain whether or not mutant-Purple would still express the purple color: we simply transformed the plasmid containing the mutant gene and observed the resulting cells for color changes.

Materials & Methods

A copy of Purple was obtained from Dr. Mikhail V. Matz (University of Texas at Austin). This copy of Purple had been inserted into both the pET16b and pET22b plasmids, which contain the gene for ampicillin resistance. Additionally, a His-tag had been added to the C-terminal of Purple to allow for purification. (See Appendix for the sequence of Purple.)

Site-Directed Mutagenesis

PCR: Polymerase Chain Reaction (PCR) was used to perform site-directed mutagenesis on the Purple gene. Each of the 3 sets of mutagenesis primers was used sequentially to change a single amino acid. (See appendix for primer sequences.) This was performed in 50μL reactions consisting of 100ng of template DNA. The template for the first round of PCR reactions was Purple in either the pET16b or pET22b plasmid. Single mutants were then used as templates to create double mutants, and the double mutants were used to create the triple mutant. The PCR reaction included: 1X reaction buffer, 2.5 units Pfu DNA polymerase, 0.25 mM dNTPs, 125ng forward primer, and 125ng reverse primer. PCR cycling was for 1 minute at 95°C and then 18 cycles of: 95°C for 50 seconds, annealing at 57.6°C (a gradient of temperatures consisting of 68°C, 63.3°C, 57.6°C, and 55.0°C was used if 57.6°C failed to produce a significant amount of product) for 50 seconds, and polymerization at 70°C for 9.5 minutes. These 18 cycles were followed by 7 minutes at 70°C and then held at 4°C.

Digestion: After the PCR reaction was completed, template DNA was digested by adding 10 units of the restriction enzyme DpnI. This was then incubated in a 37°C water bath for 90 minutes.

Analytical Gel: To confirm that the plasmid had been amplified, 10μL of each reaction was added to 2μL 5x loading buffer and electrophoresed through a 1% analytical agarose gel containing ethidium bromide for 60 minutes at 100V in 1X Tris-Acetate-EDTA buffer.

Cloning and Purification

Transformation into XL-10 Gold: Each PCR reaction was transformed into XL-10 Gold competent cells. This was done by adding 15μL of the PCR reaction to 100μL thawed, competent cells and incubating on ice for 30 minutes. The transformation was plated on 2XYT plates containing ampicillin and incubated at 37°C overnight.

Miniprep: Isolated colonies were then selected and grown in 3 mL 2XYT cultures (containing 0.15 mg of ampicillin per 1 mL of culture) and grown overnight in a shaking incubator at 37°C. These cultures were then purified to extract the pET16b or pET22b plasmid using the QIAGEN Spin Miniprep Kit following the manufacturer's instructions and eluting with 75 μL of elution buffer.

Digestion: A diagnostic digest was then performed to confirm that the proper plasmid resulted from the PCR mutagenesis. The digest was performed with 3 μL of each plasmid miniprep in 10 μL reactions containing 3 units of restriction enzyme PvuII in 1X

NEB Buffer 2 and ddH_2O. This digest was then incubated in a water bath at 37°C for 1 hour.

Analytical Gel: All 10 μL of the digestion were then added to 2μL 5x loading buffer and electrophoresed through a 1% analytical agarose gel containing ethidium bromide for 60 minutes at 100V in 1X Tris-Acetate-EDTA buffer. DNA concentrations of the plasmids were determined using a Nanodrop device.

Sequencing of Purple Mutants

To confirm that the desired mutation had been successfully made, 500ng of each mutant plasmid was sequenced by the University of Texas ICMB sequencing core facility. If the correct mutation was found to be in Purple, then this mutant plasmid was then used as the template in the next PCR reaction to add additional mutations. This cycle was repeated until all 3 mutations had been inserted to produce the triple-mutant of Purple.

Assessment of Purple Color

Transformation into BL21: After successfully creating the triple mutant of Purple, we then had to assess each of the mutant's ability to produce color. Each mutant of Purple was transformed into BL21 cells by adding 1 ng of the mutant Purple plasmid to 25 μL BL21 cells and incubating for 30 minutes on ice. This was then plated on 2XYT plates containing ampicillin and incubated overnight at 37°C. This produced a set of plates containing each mutant in BL21 cells, which could then be used to inoculate subsequent cultures.

Cultures of BL21: Isolated colonies of the triple mutants were then used to inoculate 5 identical 3 mL cultures of 2XYT (containing 0.15 mg of ampicillin per 1 mL of culture). These cultures were then incubated overnight at 37°C in a shaking incubator. Additionally, 5 identical 3 mL cultures of 2XYT (containing 0.15mg of ampicillin per 1 mL of culture) were also inoculated with the un-mutated Purple to serve as a positive control. These cultures were then diluted to an OD of 0.1, as measured by a spectrophotometer, in 10 mL 2XYT (containing 0.15mg of ampicillin per 1 mL of culture). This 10 mL set of cultures was then incubated at 37°C in a shaking incubator until they reached an OD of 0.45. At this point, the 5 cultures of both the triple mutant and the Purple control were induced to produce Purple with varying concentrations of IPTG: 0.05mM, 0.10mM, 0.25mM, 0.50mM, and 1.00mM. The cultures were then incubated at 37°C in a lighted, shaking incubator overnight and assessed for purple coloration by visual examination the next day.

Protein Purification

Cultures: 3 mL cultures of 2XYT (containing 0.15 mg of ampicillin per 1 mL of culture) were inoculated with the single mutants of Purple and the triple-mutant of Purple. A culture of the un-mutated copy of Purple was also grown. These cultures were grown at 37°C in a shaking incubator overnight. 2 ml of each culture were then used to inoculate 200 ml cultures of 2XYT (containing 0.15 mg of ampicillin per 1 mL of culture) and then grown at 37°C in a shaking incubator until an OD of 0.3 was reached. At this point, each culture was induced with a concentration 0.05 mM IPTG and then incubated at 37°C in a lighted, shaking incubator overnight. The 200 mL cultures were centrifuged at 5000 rpm for 10 minutes to pellet the cells. Pellets were re-suspended in 15mL of 20

mM $NaPO_4$ (pH7), 200 mM NaCl solution. Re-suspended pellets then underwent 4 rounds of freezing in liquid nitrogen followed by thawing in a 2°C water bath. Cells were then subjected to sonification for 5 cycles involving 15 seconds of sonification at power level 7, and 45 seconds cooling in ice water in between cycles. After sonification, 1.5 mL of 10% Triton X-100 was added to each sample and then centrifuged at 15,000rpm for 20 minutes. Pellets were discarded.

Ni-NTA Purification: The His-tag allowed us to use Ni-NTA purification to capture the protein. The supernatant containing the soluble Purple protein was run through BioRad Poly-Prep Chromatography Columns. Columns had been prepared by adding 200 µl NiNTA (beads) to each column followed by 10mL of solution (20 mM $NaPO_4$ pH 7.0, 200mM NaCl, and 10 mM β-mercaptoethanol). After chromatography, each column was washed 3 times with 10 mL of wash buffer (20mM $NaPO_4$ pH7, 200 mM NaCl, 10 mM β-mercaptoethanol, 25 mM imidazole). Each column was then eluted 5 times with 200 µL elution buffer (20 mM $NaPO_4$ pH 7.0 200 mM NaCl, 300 mM imidazole).

Gel Analysis: Samples were then prepared by adding 20 µL 2X LDS buffer to 2µL of the whole-cell sample, 2 µL of the supernatant sample, and 2 µL of the flow-through sample from each mutant. 4 µL of these samples were electrophoresed on an SDS-PAGE gel. 5µL of the 2X LDS buffer was added to 5µL of elution fractions 2 and 3 from each mutant, and this was electrophoresed on the SDS-PAGE gel. Invitrogen BenchMark Protein Ladder was used as a size standard. The protein gel was run in 1X MES SDS Page buffer at 200 V for 41 minutes and fixed for 15 minutes in 50% MeOH, 12% HoAc

solution on a shaker. After fixing the gel, it was then washed twice for 5 minutes in d_2H_2O on the shaker prior to staining with Coomassie Blue G250.

Sephadex S200 Gel Filtration Column

The Sephadex S200 Column on a BioRad Biologic LP chromatography system was used to separate the purified proteins by relative molecular weight. A total of sixty 2.5 ml elution fractions were collected from each sample.

Dot Blot

Nitrocellulose paper was prepared for the dot blot by soaking in water for 5 minutes, followed by soaking in TBS solution (20mM Tris pH 7.5, 150mM NaCl) for 5 minutes. The membrane was then inserted into a minifold apparatus, and 250 μL of every 3rd fraction of each sample was run through under vacuum suction. 10μL of the whole cell sample of Purple from the Ni-NTA purification was loaded into the last well to serve as a control. Preparation and visualization was done according to the instructions for chemiluminescent substrates in the QIAexpress Detection and Assay Handbook.

Results

Site-Directed Mutagenesis: pET16b

Mutants were first created using Purple in a pET16b vector. Thus, we obtained a number of different single, double, and triple mutants (See Table 1). Some of the mutations required slightly different annealing temperatures in the PCR reaction. Note that on each figure, mutation 1 refers to amino acid 102, mutation 2 refers to amino acid 149, and mutation 3 refers to amino acid 158.

Table 1. PCR reaction conditions for each mutant in pET16b.

Vector	Mutant	Annealing Temperature (°C)
pet16b	2	56.7
pet16b	1.2	56.7
pet16b	2.3	56.7
pet16b	1.2.3	68.0

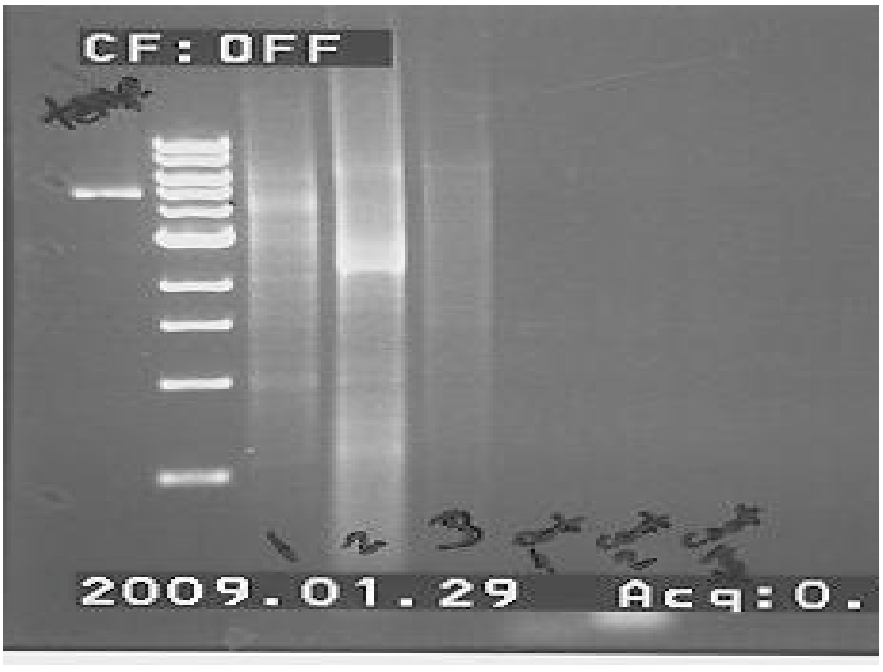


Fig 7. PCR with each single mutant in pET16b at 57.6℃ annealing temperature. From the left: pET16b template, 7.5μL 1KB ladder, mutation 1, mutation 2, mutation 3, and controls for mutation 1, 2, and 3.

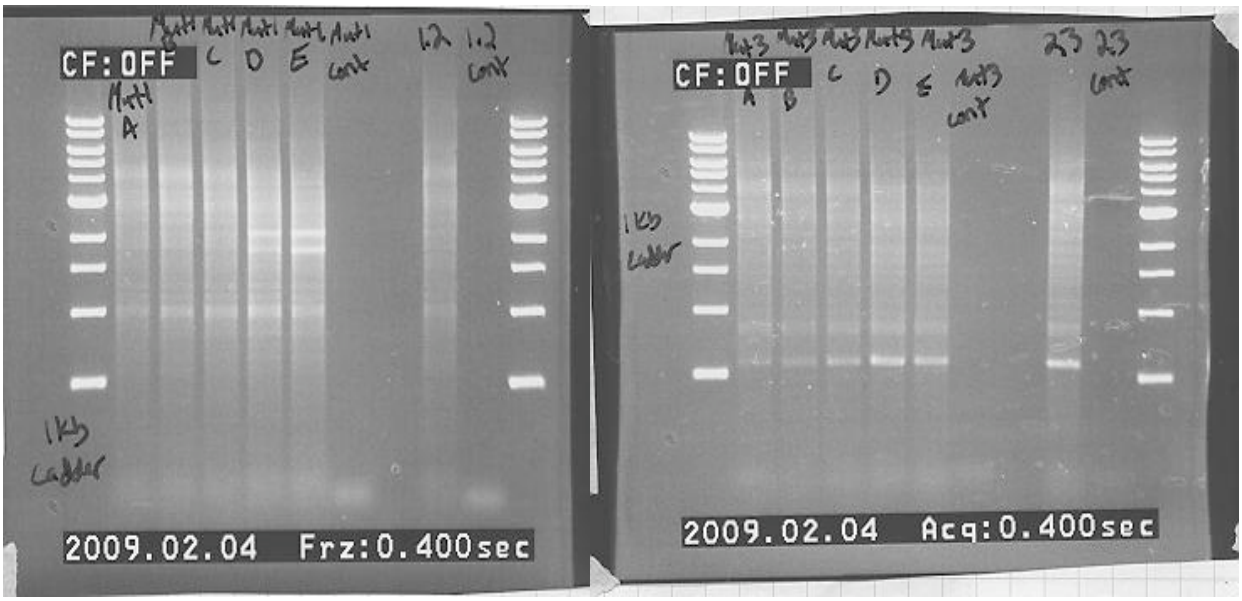


Fig 8. PCR of mutations 1, 1.2, 3, and 2.3 in pET16b. Far left and right on each gel is 7.5μL of 1 Kb ladder. Mutations from left to right include mutation 1 (gradient of temperatures: 45℃, 49℃, 57.6℃, 61.2℃, 65℃ and control), mutation 1.2 (and control), mutation 3 (gradient of temperatures: 45℃, 49℃, 57.6℃, 61.2℃, 65℃ and control), and mutation 2.3 (and control).

Additionally, we originally tried PCR with both Taq DNA polymerase, Pfu DNA polymerase, or a mixture of the 2 enzymes, and found that Pfu DNA polymerase alone worked the best (See Fig 9). Thus all PCR reactions were done using the polymerase Pfu DNA polymerase.

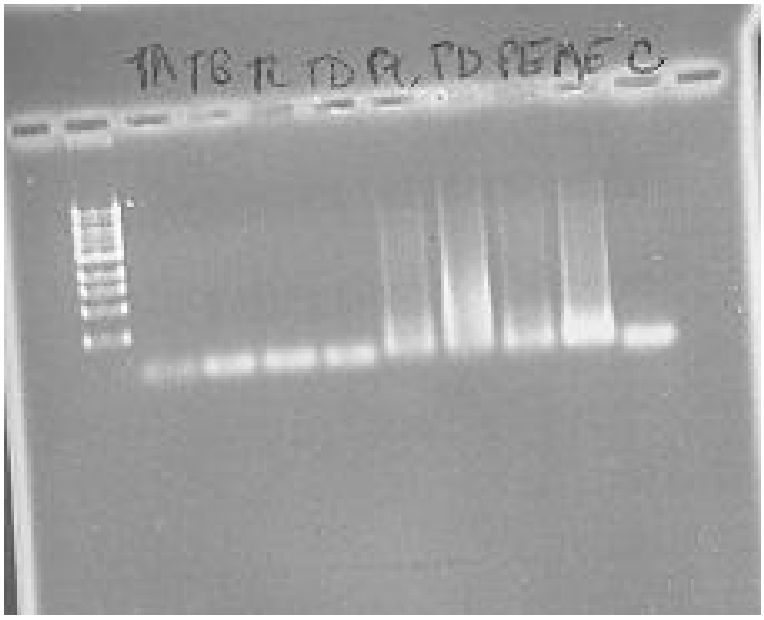


Fig 9 (Left). Gradient of PCR with a single mutation using Taq, Pfu, and a mixture of both. Left to right: 7.5μL 1Kb Ladder, Taq (gradient of temperatures: 65℃, 61.2℃, 57.6℃, 49℃, 45℃), Pfu (gradient of temperatures: 57.6℃, 49℃, 45℃), Mix 45℃, and control 57.6℃.

After transforming and cloning the PCR for each mutant, the resulting cloned DNA was digested with restriction enzyme PvuII and run on an agarose gel. This allowed us to determine if the proper bands were present (see Fig 10). However, this technique only helped to identify if the clone contained the pET plasmid, but did not indicate if the proper mutation had been made. Clones that displayed the correct bands on the gel were thus sequenced to see if they contained the correct mutation. The sequence of Purple with each mutation added can be found in the appendix. We were only able to generate one single mutant, mutant 1. However, we used mutant 1 to generate double mutants 1.2 and 2.3. We then used these to create the triple mutant.

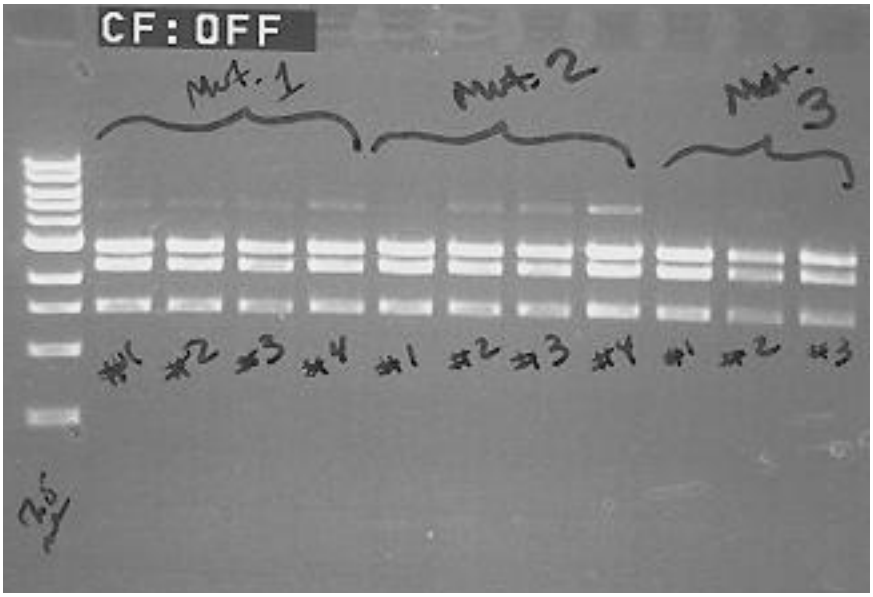


Fig 10. PvuII digestion of mutant 1, 2, and 3 clones. Left to right: 7.5 µL 1 Kb ladder, 4 clones of mutant 1, 4 clones of mutant 2, and 4 clones of mutant 3.

After generating the triple mutant, we then tried to transform each mutant into cells lines to express Purple. The first cell line we chose was Rosetta Blue DE3. However, when the mutants and control were transformed and plated onto plates containing IPTG, the

cells failed to grow. We thought it might be the case that the IPTG, which was meant to induce the expression of Purple, was preventing the cells from growing. To rule out this possibility, we grew cultures of the control and mutant, which had been transformed into Rosetta Blue DE3, to an OD of 0.3 and then induced these with 0.5 mM IPTG. These cells failed to express Purple. This caused us to re-evaluate the compatibility of Purple and the pET16b vector, and thus we chose to recreate the mutants of Purple in a pET22b vector.

Site-Directed Mutagenesis: pET22b

Mutants were generated in the pET22b vector using the same method used for the pET16b vector. Thus, we were able to generate a complete set of single mutants (1, 2, 3), 2 double mutants (1.2, 1.3) and the triple mutant (1.2.3). See table 2 for reaction conditions below.

Table 2. PCR reaction conditions for mutants in pET22bb.

Vector	Mutant	Annealing Temperature (°C)
pet22bb	1	56.7
pet22bb	2	56.7
pet22bb	3	56.7
pet22bb	1.2	56.7
pet22bb	1.3	56.7
pet22bb	1.2.3	56.7

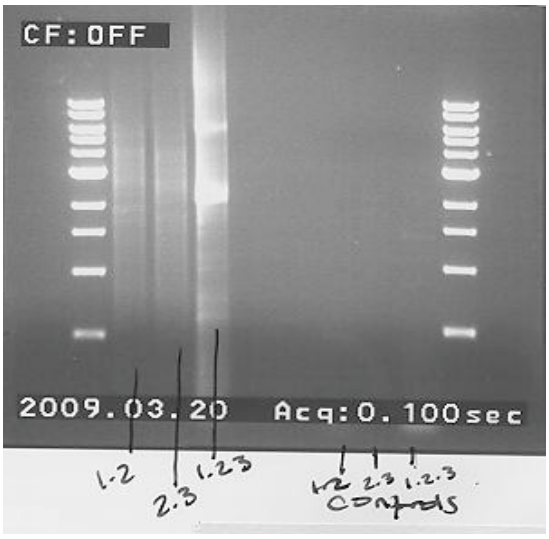


Figure 11. PCR of mutants in pET22b. Left to right: 7.5μL 1 KB ladder, mutant 1.2, mutant 2.3, mutant 1.2.3, control: mutant 1.2, control: mutant 1.3, control: mutant 1.2.3, 7.5μL 1 KB ladder.

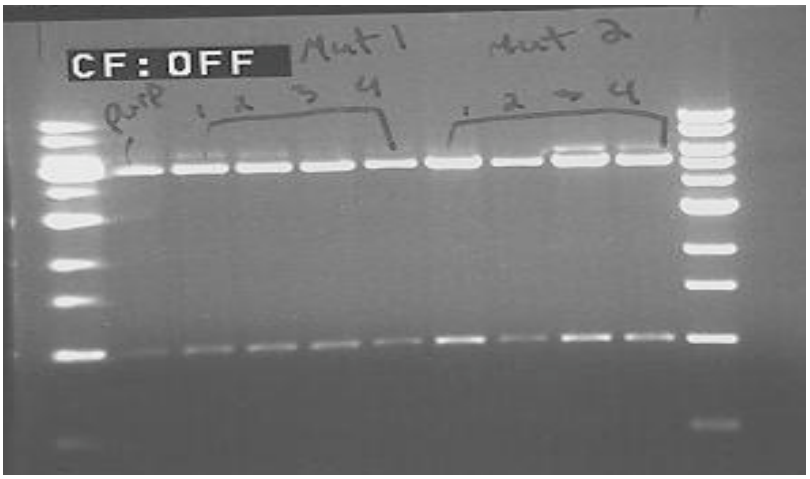


Figure 12. PvuII digest of mutant clones in pET22b. Left to right: 7.5μL 1 KB ladder, Purple in pET22b, 4 clones of mutant 1, 4 clones of mutant 2, 1 KB ladder.

After creating the set of mutations in pET22b, we then transformed all of the mutations from both plasmids (pET16b and pET22b) into BL21 cells. It turned out that Purple was strain specific, meaning that we had not been able to express it in the Rosetta Blue DE3 cells because they were the wrong strain, rather than because the plasmid (pET16b) was wrong. In order to determine the concentrations of IPTG necessary for the cells to express purple, we grew cultures of both the Purple control and the triple mutant to an OD of 0.45, and then induced them with different concentrations of IPTG: 0.05 mM, 0.10 mM, 0.25 mM, 0.50 mM, 1.00 mM. After letting the cultures continue to grow overnight, we found that the Purple control expressed the purple color at all concentrations, and the triple mutant did not turn purple at any concentration.

Protein Purification

Mutant 2 consistently failed to grow in culture, and thus was not purified. However, cultures of mutant 1, mutant 3, and the triple mutant were purified. After separating out the protein using Ni-NTA purification, we could visibly see purple coloration in the fractions of the control (un-mutated Purple), as well as mutant 1. Gels were run containing different fractions of each sample, including the whole cell (prior to

centrifugation), soluble protein (supernatant after centrifugation), flow through (protein from the sample that did not remain on the column), and fractions 2 and 3 of the elutions. These gels confirmed that there were strong and clearly identifiable bands in the eluted fractions of the mutants that matched the bands in the control Purple.

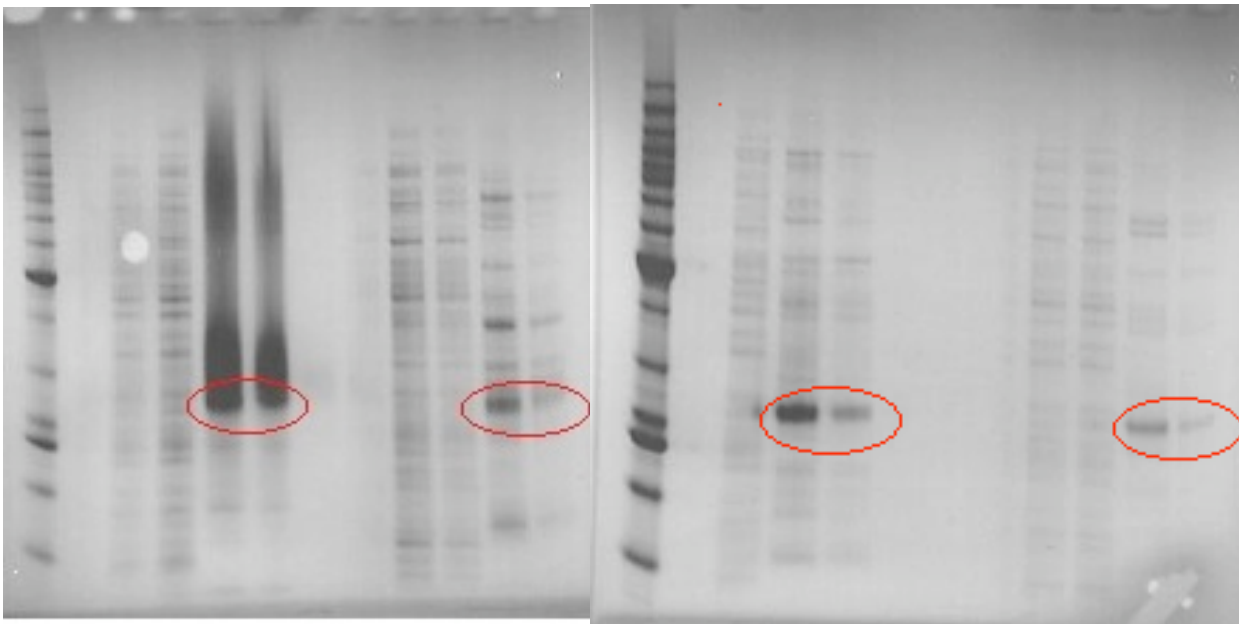


Fig. 13. Protein gel of fractions from Ni-NTA purification. Each set of mutants contains samples, (left to right) of the whole cell, soluble, flow through, fraction 2, and fraction 3. The protein band of interest has been circled (red).
Left gel: Invitrogen BenchMark Protein Ladder, Purple control, mutant 1.
Right gel: Invitrogen BenchMark Protein Ladder, mutant 3, triple mutant.

By combining data from the Sephadex gel filtration column and the dot blot test, we determined that both the control Purple and mutant 1 eluted at ~50.0mL. The dot blot of Purple and mutant 1 confirm that they both eluted in the same fraction (Fig 15: circled green). The second peak on the dot blot of the control Purple, which did not show up on the graph from the gel filtration column, is most likely degraded protein as it corresponds to the size of small peptides (Fig 15: circled blue). By comparison with a size standard, 50.0 mL corresponds to a size of 65kD. By contrast, mutant 3 eluted at

~64.0 mL, which corresponds to ~32kD (Fig 15: circled yellow). Unfortunately, there was insufficient material for the gel filtration column of the triple mutant to give a clear peak. This was confirmed by the dot blot, which indicated that there were no fractions of the triple mutant strong enough to be distinguished from background.

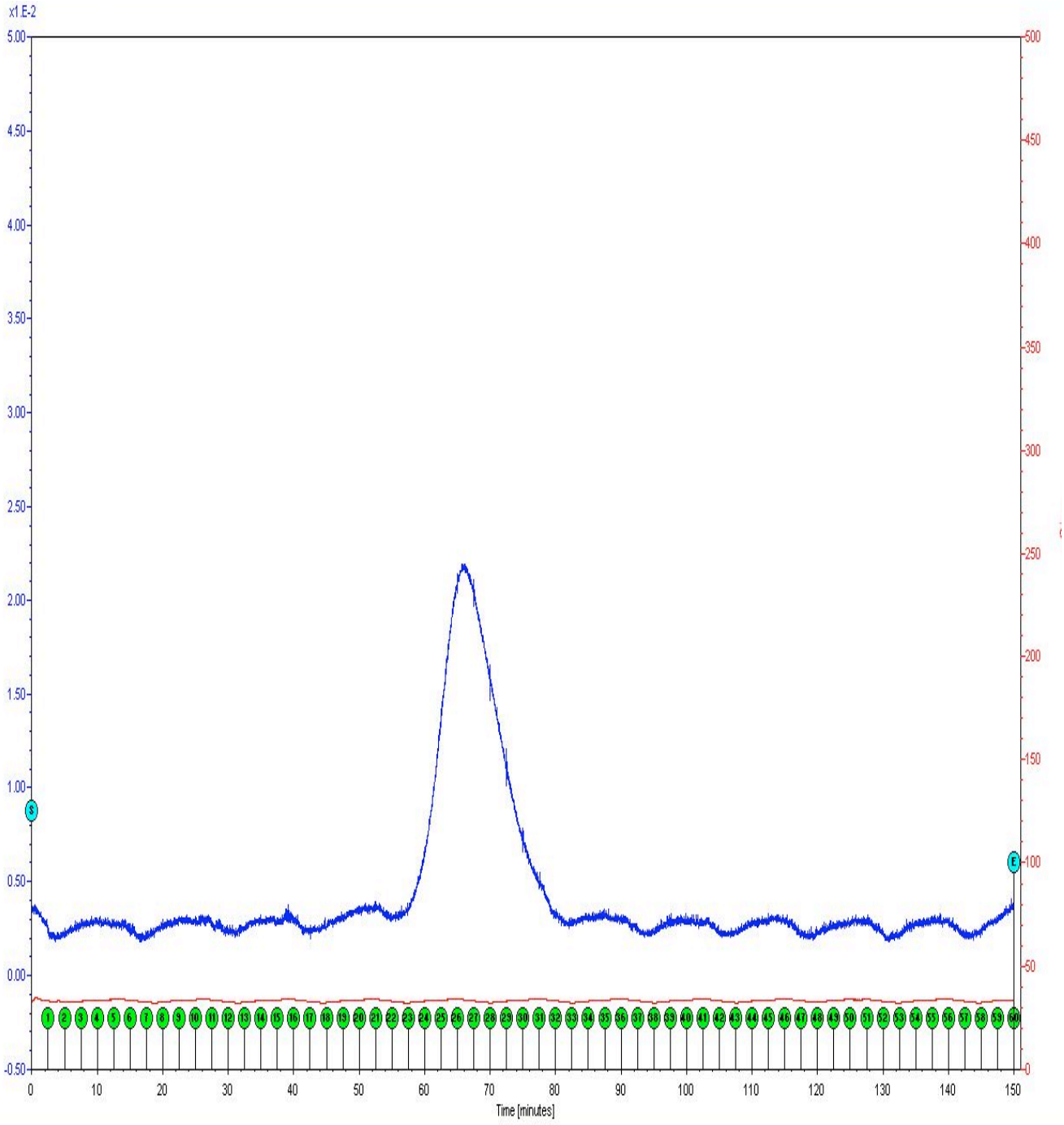


Fig. 14. Elution peak of Purple control on the Sephadex S-200 gel filtration column.

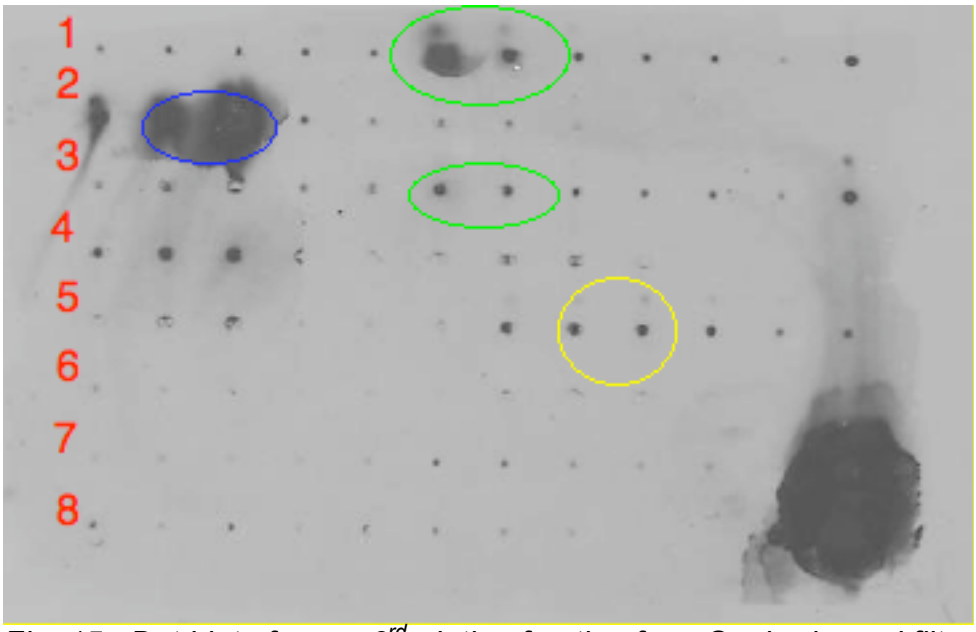


Fig. 15. Dot blot of every 3rd elution fraction from Sephadex gel filtration.
Row 1-2: Purple control.
Row 3-4: Mutant 1.
Row 5-6: Mutant 3.
Row 7-8: Triple Mutant.
Last well (bottom right): Purple control, whole cell fraction from Ni-NTA purification.

Discussion

In summary, this study focuses on a member of the GFP-family, Purple protein, recently discovered by Dr. Mikhail V. Matz. Very little is known about the 3-dimensional structure of the protein. However, it is commonly believed that almost all members of the GFP-family naturally occur as oligomers, with most being found as tetramers. Because we had no reason to believe that Purple would not follow the same pattern as similar GFP-like proteins, in addition to early experiments conducted by Dr. Matz, we assumed that Purple would also form tetramers. This is an undesirable property for Purple to have, in terms of the possible biotechnology applications of the protein.

Thus, the goal of this study was to monomerize purple via site-directed mutagenesis of 3 amino acids. Additionally, it was our hope that the monomerized Purple protein would retain its ability to produce the color purple when expressed. Based on sequence similarities to a very similar protein, amino acids 102, 149, and 158 were chosen as mutations that were likely to produce monomeric Purple.

After analyzing the resulting mutants, the data indicates that mutating amino acid 102 did not affect the oligomerization state of Purple. However, it clearly affected the spectral properties of Purple, as it was no longer able to produce purple coloration. Interestingly, the 65kD relative molecular mass that characterizes both Purple and the version of Purple with a mutation in residue 102 indicates that Purple might actually occur naturally as a dimer, rather than as a tetramer as was previously thought. We could not determine the oligomerization state of Purple with the mutation of amino acid

149, because this mutant repeatedly failed to grow in cultures. However, based on that piece of information alone, it is unlikely that this mutation would be successful in producing a stable monomer that still expressed Purple. Finally, by mutating amino acid 158, the oligomerization state of Purple seems to have been changed. A size of 32kD indicates that this mutant is half the size of wild-type Purple.

If we believe that Purple is indeed a dimer, then we have successfully monomerized Purple with this mutation. However, even with the more conservative assumption that Purple is naturally a tetramer, we still have successfully changed the oligomerization state of Purple with this mutation, which is a very useful piece of information to make future alterations to the protein. Unfortunately, this mutation has affected the spectral properties of Purple, effectively knocking out its ability to produce color. While we were not able to draw any conclusion as to the oligomerization state of the triple-mutant from the purification and column chromatography experiments, it is likely that this mutant has similar properties to the version with the mutation in residue 158. It is possible, however, that the expression of Purple in the triple mutant is hindered by the mutation in residue 149, which seems to have adversely effected the single-mutant containing that particular mutation (mutant 2).

Future Directions: The results of this study have established a clear groundwork for future experiments on Purple. For one, it would be desirable to determine the wild-type oligomerization state of Purple, through such techniques as analytical ultracentifugation. This would help to definitively ascertain if mutating residue 158 has produced a

monomer or a dimer. Additionally, now that the oligomerization state of Purple has been altered, it may be possible to recover the original spectral properties of the protein by performing mutational screens of this mutant. In this way, the original goal of creating a monomeric Purple protein may still be achieved, despite the fact that the mutation, which monomerized Purple, disrupted its ability to produce color.

Appendix

Primer Sequences. Note: lower case letters represent mutations.

Mutant 1	T102K - Forward	GAAGATGGTGCAGTGTGTaaaGTCAGCAATGATTCCAGC
Mutant 1	T102K – Reverse	GCTGGAATCATTGCTGACTTTACACACTGCACCATCTTC
Mutant 2	R149E – Forward	ACTGAGCGTCTCTTTGCAgaaGATGGAATGCTGATAGGA
Mutant 2	R149E – Reverse	TCCTATCAGCATTCCATCTTCTGCAAAGAGACGCTCAGT
Mutant 3	F158K - Forward	ACTGAGCGTCTCTTTGCAgaaGATGGAATGCTGATAGGA
Mutant 3	F158K - Reverse	TTCCAACTTCAGAGCCATTTTGTTGTTTCCTATCAGCAT

Sequence of Purple (unmutated):

ATGTCAGGCACGGTCAATGGACACTACTTTGAGGTCGAAGGCGATGGAAAAGGAA
AGCCTTACGAGGGGGAGCAGACGGTAAAGCTCACTGTCACCAAGGGCGGACCTC
TGCCATTTGCCTGGGATATTTTATCACCACAGTCTCAGTACGGAAGCATACCATTCA
CCAAGTACCCTGAAGACATCCCTGACTATGTAAAGCAGTCATTTCCTGAGGGATAT
ACATGGGAGAGGATCATGAACTTTGAAGATGGTGCAGTGTGTACTGTCAGCAATGA
TTCCAGCATCCAAGGCAACTGTTTCATCTACCATGTCAAGTTCTCTGGTTTGAACTT
TCCTCCCAATGGACCTGTTATGCAGAAGAAGACACAGGGCTGGGAACCCAACACT
GAGCGTCTCTTTGCACGAGATGGAATGCTGATAGGAAACAACTTTATGGCTCTGAA
GTTGGAAGGAGGTGGTCACTATTTGTGTGAATTCAAATCTACTTACAAGGCAAAGA
AGCCTGTGAAGATGCCAGGGTATCACTATGTTGACCGCAAACTGGATGTAACCAAT
CACAACAAGGATTACACTTCCGTTGAGCAGTGTGAAATTTCCATTGCACGCAAATC
TGTGGTCGCCCATCACCATCACCATCATCATGAAAACCTGTATTTTCAGCAT

Sequence of Purple triple-mutant:

ATGTCAGGCACGGTCAATGGACACTACTTTGAGGTCGAAGGCGATGGAAAAGGAA
AGCCTTACGAGGGGGAGCAGACGGTAAAGCTCACTGTCACCAAGGGCGGACCTC
TGCCATTTGCCTGGGATATTTTATCACCACAGTCTCAGTACGGAAGCATACCATTCA
CCAAGTACCCTGAAGACATCCCTGACTATGTAAAGCAGTCATTTCCTGAGGGATAT
ACATGGGAGAGGATCATGAACTTTGAAGATGGTGCAGTGTGTaaaGTCAGCAATGA
TTCCAGCATCCAAGGCAACTGTTTCATCTACCATGTCAAGTTCTCTGGTTTGAACTT
TCCTCCCAATGGACCTGTTATGCAGAAGAAGACACAGGGCTGGGAACCCAACACT
GAGCGTCTCTTTGCAgaaGATGGAATGCTGATAGGAAACAACaaaATGGCTCTGAAG
TTGGAAGGAGGTGGTCACTATTTGTGTGAATTCAAATCTACTTACAAGGCAAAGAA
GCCTGTGAAGATGCCAGGGTATCACTATGTTGACCGCAAACTGGATGTAACCAATC
ACAACAAGGATTACACTTCCGTTGAGCAGTGTGAAATTTCCATTGCACGCAAATCT
GTGGTCGCCCATCACCATCACCATCATCATGAAAACCTGTATTTTCAGCAT

Works Cited

Alieva, N.O., et al. Diversity and Evolutions of Coral Fluorescent Proteins, *Plos One*. **3**: e2680 1-12. 2008.

Baird, G.S., D.A Zacharias, and R.Y. Tsien. Biochemistry, Mutagenesis, and Oligomerization of DsRed, A Red Fluorescent Protein From Coral, *PNAS*. **92:** 11984-11989. 2000.

Delagrave, S., et al. Red-Shifted Excitation Mutants of the Green Fluorescent Protein, *Nature Biotechnology*. **13:** 151-154. 1994.

Lippincott-Schwartz, J., et al. Development and Use of Living Protein Markers in Living Cells, *Science*. **300**: 87-91. 2003.

Lukyanov, K.A., et al. Natural Animal Coloration Can Be Determined by a Nonfluorescent Green Fluorescent Protein Homology, *The Journal of Biological Chemistry*. **275**: 25879-25882. 2000.

Matz, M.V., K.A. Konstantin, and S.A. Lukyanov. Family of the Green Fluorescent Protein: Journey to the End of the Rainbow, *BioEssays*. **24**: 953-959. 2002.

Misteli, T. and D.L. Spector. Applications of the Green Fluorescent Protein in Cell Biology and Biotechnology, *Nature Biotechnology.* **15:** 961-964. 1997.

Ormo, M., et al. Crystal Structure of the *Aequorea victoria* Green Fluorescent Protein, *Science*. **273**: 1392-1395. 1996.

Pakhomov, A.A. and V.I. Martynov. GFP Family: Structural Insights into Spectral Tuning, *Chemistry & Biology Review*. **15**: 755-764. 2008.

Phillips, G.J. MiniReview: Green Fluorescent Protein – A Bright Idea for the Study of Bacterial Protein Localization, *FEMS Microbiology Letters*. **204**, 9-18. 2001.

Shaner, N.C., P.A. Steinbach, and R.Y. Tsien. A Guide to Choosing Fluorescent Proteins, *Nature Methods* **2:** 905-905. 2005.

Tsien, R. J. The Green Fluorescent Protein, *Annu. Rev. Biochem*. **67**: 509-544. 1998.

Yang, F., L.G. Moss, and G.N. Phillips. The Molecular Structure of Green Fluorescent Protein, *Nature Biotechnology*. **14**: 1246-1251. 1996.